

Use of Electronic Health Records to Support a Public Health Response to the COVID-19 Pandemic in the United States: A Perspective from Fifteen Academic Medical Centers

Madhavan S¹, Bastarache L², Brown JS³, Butte A⁴, Dorr D⁵, Embi PJ⁶, Friedman CP⁷, Johnson KB⁸, Moore JH⁹, Kohane IS¹⁰, Payne PRO¹¹, Tenenbaum JD¹², Weiner MW¹³, Wilcox A¹⁴, Ohno-Machado L^{15,16#}

¹ Innovation Center for Biomedical Informatics, Georgetown University Medical Center Washington, DC

² Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN

³ Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, MA

⁴ University of California Health System (UC Health), University of California, San Francisco, CA

⁵ Departments of Medical Informatics and Clinical Epidemiology and Medicine, Oregon Health and Science University, Portland, OR

⁶ Indiana University School of Medicine, Regenstrief Institute, Inc., Indianapolis, IN

⁷ Department of Learning Health Sciences, University of Michigan, Ann Arbor, MI

⁸ Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN

⁹ Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania Philadelphia, PA

¹⁰ Department of Biomedical Informatics, Harvard Medical School, Boston, MA

¹¹ Institute for Informatics, Washington University in St. Louis, School of Medicine, St. Louis, MO

¹² North Carolina Department of Health and Human Services, Raleigh, NC

¹³ Department of Population Health Sciences, Weill Cornell Medicine, New York, NY

¹⁴ Biomedical Informatics and Medical Education, University of Washington, Seattle, WA

¹⁵ Department of Biomedical Informatics, Associate Dean for Informatics and Technology, University of California San Diego Health, La Jolla, CA

¹⁶ Division of Health Services Research & Development, VA San Diego Healthcare System, San Diego, CA

#Corresponding Author

Lucila Ohno-Machado, MD, PhD

Professor of Medicine

Chair, UC San Diego Health, Department of Biomedical Informatics

Associate Dean for Informatics and Technology

University of California School of Medicine

lohnomachado@health.ucsd.edu

Abstract

Our goal is to summarize the collective experience of fifteen organizations in dealing with uncoordinated efforts that result in unnecessary delays in understanding, predicting, preparing for, containing, and mitigating the COVID-19 pandemic in the USA. Response efforts involve the collection and analysis of data corresponding to healthcare organizations, public health departments, socioeconomic indicators, as well as additional signals collected directly from individuals and communities. We focused on Electronic Health Records (EHRs) data, since EHRs can be leveraged and scaled to improve clinical care, research, and to inform public health decision-making. We outline the current challenges in the data ecosystem and the technology infrastructure that are relevant to COVID-19, as witnessed in our fifteen institutions. The infrastructure includes registries and clinical data networks to support population-level analyses. We propose a specific set of strategic next steps to increase interoperability, overall organization and efficiencies.

Introduction

Responding to a public health emergency, such as the current COVID-19 pandemic, requires access to and analysis of large and timely volumes of data[1] with goals of providing timely and appropriate information to policy makers, keeping the public safe by making resources available where needed, and conducting research to advance our collective knowledge of such pandemics for future preparedness. Unfortunately, there are many barriers in building a robust and scalable infrastructure needed to enable such “digital public health”[2]. Investments of federal, state, county, and city funds, as well as of private funds are limited by political and geographic divisions[2]. When emergent situations require rapid response by public health departments, and extant infrastructure and data are not sufficient to meet those needs, the solution often involves the use of an emergency public health directive whereby healthcare provider organizations (HPOs) and supporting clinical services (e.g., clinical laboratories, etc.) are required to submit data for reporting purposes[3] using a myriad of often changing data definitions, encoding standards, and transmission mechanisms -- all of which incur substantial burden to those organizations[4]. The lack of coordination between and among various entities, as well as the insufficiency of technical infrastructure[5], result in substantial expense, sub-optimal outcomes, and missed opportunities for timely and impactful public health interventions[6]. Here, we address the current challenges and limitations and propose realistic steps towards a health information infrastructure that serves the needs of patients, clinicians and public health authorities to address pandemics in a timely fashion.

EHRs contain many important data elements that can help with a pandemic response[7,8]. Although EHRs have known shortcomings as the sole source of data for studies that inform public health decisions[9], utilization of a large number of records from many institutions could help provide mission-critical answers to clinicians, researchers, administrators, public health officials and the public in general[10].

Limitations of current initiatives to use EHRs beyond the care of an individual

Multiple initiatives now exist to physically or virtually aggregate EHR and related data for COVID-19[11,12], and single system data have been used to support clinical investigations of a few health conditions[13]. These efforts in the USA are in large part limited to academic medical center networks and have not been focused on public health, although some initial actions hold promise. For example, specific codes for COVID-19-related data have been introduced into controlled terminologies employed in EHRs or their derivatives; large consortia are trying to build robust cohort definitions to ensure reproducibility in the harmonization and utilization of concepts across various institutions. In the USA, many EHR-based initiatives consist of building (a) specific registries for SARS-COV-2 tested individuals and for those diagnosed with COVID-19, or (b) activation of clinical data networks (CDNs) to access COVID-19 data included in EHRs. **Table 1** shows some differences between these types of initiatives.

Academic medical centers have been responsible for a large portion of research involving EHRs[14,15]. From the perspective of our 15 institutions and from interacting with colleagues in similar positions, EHR-based COVID-19 data collection and sharing initiatives are currently stretching the roles of information technology (IT) and informatics teams within the health systems providing such data, which are also trying to provide care, conduct research and educate healthcare workers during the pandemic. Registries and CDNs frequently adopt different data models, standards, and governance structures, leading the Food and Drug Administration (FDA) to sponsor a project to “harmonize a list of COVID-19 data elements with several Common Data Models (CDMs) and open standards”[16]. In our experience, although many HPOs participate in more than one initiative, few resources have been added to IT and informatics teams in order to support COVID-19 specific activities, and few existing responsibilities have been scaled back. While a larger survey is needed in order to represent the perspective of a larger spectrum of medical centers, the COVID-19 pandemic has revealed important challenges in responding to EHR data-related requests in an accurate and timely fashion at our institutions. Some are related to the nature of the data; for example, any inferences on therapies through data aggregation across institutions must take into account the difficulties in accounting for important confounders related to practice variations, availability of medications, patient socioeconomic status, type of insurance coverage and other biases. Thus EHR-based evidence must not be interpreted as a substitute for well-designed randomized trials[17]. On the other hand, EHRs can provide an important resource to describe outcomes or interventions in real life, since strict eligibility criteria and design constraints do not always allow extrapolation of clinical trial findings to the population at large. Utilizing EHRs in support of reporting for the COVID-19 pandemic unveiled the fragility of the health data infrastructure in support of a public health emergency, in a similar way as others have discussed for comparative effectiveness research[18] and healthcare quality improvement[19]. The focus of this perspective is not to assess the value of EHRs for observational studies, but instead to describe the inadequacy of infrastructure to use EHRs for public health reporting. We do not cover in this article the inadequacy of this infrastructure for other types of data (e.g., environmental variables, geo-location, social media, etc.).

To understand how IT teams are being stretched thin, consider the example list of COVID-19 initiatives (**Figure 1, Table 2**) that are requesting COVID-related EHR or associated data from healthcare and public health organizations to then make available de-identified individual patient-level records or aggregated data for a variety of operational, population and clinical or translational research purposes. Such COVID-19 data sharing and reporting activities (at the national level) are currently led by a combination of health agencies (e.g., CDC's National Health Safety Network, FDA's Evidence accelerator), media and IT companies (e.g. New York Times COVID-19 dashboards, Apple's mobility data), academic medical centers (e.g., Hopkins Resource, COVID-19 Data Discovery Index) and professional societies (e.g., American Society of Microbiology's COVID registry, Society of Critical Care Medicine's VIRUS registry). This "snapshot" of COVID-19 data and analytic resources spanning patient information from EHRs to social media and population-level summaries is evolving quickly, as are governmental and philanthropic funding opportunities for related efforts. We display simple axes of "public/private" access to data and whether "individual records or data aggregates" to illustrate the diversity of initiatives at a single point in time, and to show the dichotomy between data accessibility and detail or granularity. This illustration of the COVID-19 data sharing "environment" also suggests that connections among the initiatives are unclear for those who are not intimately familiar with their organization, with many initiatives investing efforts in organizing data from the same patients or populations, with ensuing deleterious outcomes in terms of costs, efficiency, and time-to-insight. Further, in our experience many institutions are maintaining different data marts for each initiative, often where different "common" data models are being utilized (thus mitigating the benefit of such "common" data models). In addition, different types of approvals for sharing patient-level and group-level (i.e., aggregate) data with various types of data "consumers" are also required, adding further regulatory and data privacy complexity. Most initial efforts should continue, but increasing attention needs to be paid to duplications and gaps, as well as to a common convergence point, which should include integration into public health information systems.

Current challenges to connect EHR and public health information systems

In addition to research-related use of data, there is also a need for the collection and integration of data from EHR systems and its communication to public health information systems in order to inform critical policy-making and intervention planning. Additionally, forecasting the need for ICU beds, ventilators, protective gear, etc., depends heavily on data generated in medical centers. While not highly dissimilar from many biomedical research uses of such data, the timeliness and nature of public health data needs does differ substantially. EHRs focus on individuals, while public health focuses on populations. While there can be shared data sources such as the EHRs, a myriad of other public health data from test sites detached from HPOs, county data, and care delivered at facilities such as nursing homes[20] and prisons does not typically get included in HPO records. The current state of COVID-19 data reflects a patchwork of uncoordinated, temporary fixes to a historically neglected public safety function. As the US enters its second decade of nationally-coordinated digital infrastructure for healthcare delivery and to modernize patient care[21], COVID-19 has demonstrated that this infrastructure is inadequate to respond to public health emergencies[22]. Developing interconnected health data nodes that include but are not limited to EHRs, public health surveillance and reporting systems, disease

registries, and patient-reported data is critical to a COVID-19 response and multiple other health conditions. An IT infrastructure to support public health that leverages EHRs and associated health data is needed, but cannot be built overnight.

While public health tools for horizon scanning, disease surveillance, epidemiological modeling, capacity planning, “hot spotting,” and targeted intervention strategies (such as isolation or contact tracing in the case of a transmissible pathogen) use as much available data as possible, the speed with which these data are collected, organized and analyzed is slow. The mode of data collection/transmission challenges aggregation, harmonization, and analysis. Also, analytic methods lag behind, and much information contained in the EHRs ends up underutilized. In our experience, there is a high burden of ad-hoc reporting mechanisms under emergency public health directives on HPOs whose primary role is to support clinical service providers. Developing a scalable solution that serves the needs of the current pandemic, leverages national investments in EHR systems, and anticipates public health data needs for the future could positively impact the management of “chronic” public health problems and stretch public dollars invested in a myriad of disconnected efforts that may not result in durable solutions. Investments that fill notable gaps between EHR systems and public health data needs could be a first step towards a coordinated effort to prepare for and respond to public health emergencies. Even when population health tools are deployed, they typically do not allow for cross-institutional data utilization. EHRs are not focused primarily on enabling population or public health responses, much less systematic or rigorous analyses that we need in order to enable the kinds of insights and responses that crises like COVID-19, the opioid crisis, and natural disasters require. A single HPO may not have sufficient numbers to discover patterns in the data, and vendor-specific population health tools may be limited to institutions using the same system. The efforts shown in **Figure 1**, for example, do not use vendor-based cross-institutional population health tools.

The public health information infrastructure does not currently support large-scale coordination[23]. In our experience, gathering information in an evolving pandemic requires multiple submissions for each case to various agencies, frequent changes to concept definitions, and multiple kinds of data; building capabilities that can facilitate efficiency requires thoughtful logic and standards. It requires a concerted initiative to overcome complex socio-cultural, ethical, legal, and trust dimensions that underpin all of the preceding issues, influencing both short and long-term decision-making, and further, the engagement of at-risk or underserved communities.

Preparation, installation, storage, maintenance, and utilization of each COVID-19 data resource incurs costs paid by various public and private institutions and internal HPO resources; the latter have been dramatically depleted due to the pandemic, with most if not all HPOs posting unprecedented losses in revenue. Even taking financial costs aside, the complex web of initiatives can be confusing, qualified personnel to operate them is scarce, and results may be suboptimal because not all important data are captured in structured format. Although many health information exchanges (HIEs) promote data exchange to support health care, they are seldom able to be used for research or public health. In their current stage, most HIEs do not retain copies of detailed individual-level data, which could be used to create a community-wide, longitudinal patient record, and therefore cannot be used for public health [22]. However, many relationships and tools developed for HIEs could be leveraged to accelerate

efforts in connecting EHR data for public health purposes. Although, ideally, longitudinal data would be collected, since measurement of outcomes such as readmissions can be undercounted if the readmission occurs at a different medical center, data from single hospitalizations can be useful to help assess clinical course and short-term, in-hospital prognoses.

A call to action

It is time to start recognizing teamwork and investing in sustainable solutions to modernize and connect public health and healthcare IT infrastructures[24], as well as framing roles for each member of the team to scale up local efforts. The US's inadequate response to the pandemic might be associated, in part, with inadequate coordination of data from clinical settings. Academic Medical Centers (AMCs) are important sources of data, but critical data cannot originate from AMCs only. It must come from other HPOs such as nursing homes, community hospitals and federally qualified healthcare centers that often are not equipped to participate in IT consortia, even though some of these have functional EHR systems. These entities cover a significant portion of the underserved population, so it is imperative that they are represented in the data resources. Even for academic medical centers, while some were quick to respond to changes dictated by the COVID-19 crisis and prepare data for internal use and sharing with other entities, others lagged behind in understanding what to do, when and how. Informal processes of consulting IT colleagues at other institutions filled an immediate gap to patch problems quickly, but this was not scalable, leaving a significant population of IT and informatics personnel solving and resolving the same problems locally.

While we do not advocate for centralization of efforts or picking one solution to fit all needs, we trust that a well-articulated national strategy and efficient, inclusive **bridges** can be built across various initiatives, and a convergence strategy must be designed. The federal government could play the same founding role as it did with the Internet and allow this infrastructure to develop and thrive independently after an initial injection of resources and buy-in of a well-designed strategy that engages multiple stakeholders. The funds may even already exist, but without coordination there will continue to be duplications and gaps, and no clear point of convergence. Key to the organization of multiple COVID-19 data resources will be a divide-and-conquer strategy, assembling cooperative teams that promote innovation to overhaul the status quo in three key components areas:

Technology

Arguably the easiest area to address, technical issues must be resolved in order to allow easy interoperability across institutions and across EHR- and public health systems. At a minimum, we need:

1. Use as few **common data models, standards, and base analytic tools** as possible to promote scientific discoveries for COVID-19 and beyond, to permit exploratory analyses and easy reproducibility of hypothesis testing. While the data may reside in different systems, a standardized way to map and consult the data across various systems should be developed. Hardened application programming interfaces (APIs) for EHRs and clinical studies should be easily available. These APIs could allow data sharing while also supporting governance requirements, including patient/institution consents, transparency in data usage, etc.

Centralization of data in limited scope registries for which governance and access rules are not established upfront should be avoided, and transparency in the oversight on use of all data should be increased. As mentioned before, while data provenance and curation are easier to verify in registries, data changes over time and researchers' need to establish benchmarks beyond the immediate focus of the registry require broader data. Use of EHR-based data in CDNs is very helpful in this context, but more difficult to coordinate.

2. Explore ways to **extend and enhance existing data interchange standards and interfaces** to ensure they can fully support the needs of public health departments in the context of an emergent situation (e.g., common data models, terminologies, APIs). These efforts should also include the engagement of HIEs and HPOs to determine how and in what capacities they can support such efforts. Existing tools should be considered before creating new ones. No new "common" data models should be created, as existing ones can be extended and, ideally, converge into fewer ones.
3. **Deploy technologies that help ethics boards** evaluate and monitor patient data use for research and public health quickly. Obsolete regulations that neither protect patient data or allow data to be used properly must be replaced.

Governance and Policy

Twentieth century regulations and practices plague efforts to share data responsibly. For example, there is little in between (a) unlimited use of data for a public health emergency by governments and their associates and (b) overly-restricting regulations on EHR data sharing by clinicians and researchers, even during a pandemic. Transferring non-consented so-called "de-identified" data or limited data sets into centralized registries is still considered more practical than solutions that keep data at HPOs. Twenty-first century artificial intelligence algorithms in which no human actually "sees" the patient-level data are left out of consideration in the privacy debate, as are emergency research initiatives. We recommend:

1. **Elimination of systematic barriers across the data ecosystem - obsolete processes** that rely on synchronous decision making by often voluntary or sometimes disengaged/uninformed bodies, certain information security officers and business executives, can cause unnecessary delays in processing data requests by Institutional Review Boards, contracting offices (data use/data sharing agreements, Business Associate Agreements with clinical partners), even in an emergency situation, despite specific directives [25], [26]. Health data that could be immediately useful for public health are not necessarily key deliverables for healthcare systems. They should be, but county, state and federal directives do not always make it clear what needs to be reported.
2. **Investigation of both centralized, decentralized and hybrid solutions** to enhance transparency into who has access to the data for what purposes, under what type of consent. This requires the implementation of a **system of accountability for using data** (i.e., rules of engagement or behavior). It requires the development of trust and verification systems, e.g., effective log analysis tools and a well specified and auditable code of conduct. Over the past decade, most healthcare organizations had to respond to breaches and penalties, at the same

time when there has been an increase in demand for electronic health information. Incentives for sharing have declined while the risks, including cyberattacks, have increased. The perceived value of sharing is relatively low by those who do not understand the importance of large sample sizes. Ongoing training is critical to ensure that people accessing the data systems are qualified to make optimal use of them. Ethics boards must continue to enforce the highest standards for COVID-19 data, with modernized processes. We do not advocate for government-controlled repositories of EHR data for COVID-19 patients, as they constitute a single point of failure as well as a single point of breach. We also do not advocate for or against commercial repositories, but we advocate instead for a system of accountability and transparency that promotes efficient, ethical use of EHR data even in public health emergencies. Involvement of law makers, researchers, clinicians, patient representatives, and government agencies will be key to the development of a new strategy. Institutions such as the National Academies or professional organizations such as AMIA should be summoned to promote open discussion and propose solutions.

3. **Creation or enhancement of coordinating bodies spanning public health departments, HPOs, and supporting clinical service providers** (e.g., laboratories), to ensure that processes are established and harmonized (in advance) for the rapid transfer and integration of data in a public health emergency that will work across a multi-EHR ecosystem. This would ideally involve the establishment of standard operating procedures and well-defined definitions of minimum data sets, transmission and encoding standards. Expectations for the frequency of data refreshes and submissions must be commensurate with the specific situation and resources.
4. **Coordination of virtual, harmonized “clearinghouses” for digital public health** at the local, regional, state, federal and global levels, wherein data flow, integration, and analytics can serve local needs and be communicated one step above or below, in such a manner that eliminates communications that skip levels and add burden, and thus increases speed, efficiency, and economies of scale. These could be transient (in an emergency) or durable (to assist in the satisfaction of chronic public health challenges), and could involve the engagement of impartial “honest brokers.” As part of such “clearinghouses”, clear expectations about data de-identification, access controls, and the “firewalling” of data where competitive or contrary interests exist, will be critical. Neutral, highly specialized auditing teams should be allowed to evaluate data, processes, and analytical results and reproduce results. Various local, state, and federal agencies must agree on how their roles and responsibilities complement each other, and make them clear to the public.

Strategy

Just as the nation has benefited from investments over the past 15 years to encourage EHR adoption and “meaningful use”, a high level of investment is needed today to ensure we are ready for future crises. Significant improvements and capabilities in recent years in EHR adoption allow us to respond to this crisis in ways that would not have been possible a decade ago. And yet, there remain inadequacies in our collective health IT infrastructure that make responding to population-level events far more challenging than they should be.

To implement the recommendations above, rather than focusing on such needs as “secondary” uses of EHRs, HIEs and the data they produce and manage, the use of our Health IT infrastructure for such activities must become another “primary” function of such systems. Achieving this requires a holistic and systematic approach (**Figure 2**) that not only focuses on technological solutions, but also focuses on the regulatory, financial and socio-organizational alignments that make such uses possible, or create head-winds when they are not addressed. Only by doing so will we create a health IT infrastructure that simultaneously allows for excellent individualized, precision healthcare, while also enabling evidence-generation and the creation of a learning health system. There needs to be a significant effort in designing a system of incentives and investments to make data sharing for public health a measurable, worthwhile component of what HPOs do. Therefore, we are calling for a bold program that capitalizes and builds upon investments to date. We should:

1. Make “Hard decisions:” While acknowledging the noble intent and value of initiatives to date, recognize that the way to move forward requires changes. For example, we need to address how data are collected, entered into EHR systems, and how these data will be used *beyond individual patient care*. We thus need to design better interfaces that **enable data to be captured according to standards** while allowing clinicians to provide care without extra barriers.
2. Provide incentives to **learn from successful HIEs, disease registries and CDNs** and develop novel models for networks of EHRs that can be used for public health and research in addition to clinical care in a manner that leverages, but is not bound or limited by proprietary, vendor-based concerns, and that also protects patient privacy. For example, require that vendor systems provide robust tools that enable data to map to common data models.
3. **Engage the whole IT, informatics, public health, epidemiologic, and administration** communities (public, private, and non-profit sectors) for common understanding of informatics issues and concerns, including standardization of epidemiologic research designs and methods that use EHR data. Address hurdles for data sharing due to commercial interests.
4. **Invest research dollars in novel, integrative methods, tools, and technology.** In parallel with the investigation of therapies and vaccines, information technology tools will be required to prioritize and organize the logistics of these health interventions and evaluation of their efficacy. While investments towards a few “data coordinating center” roles have been announced, they are scattered and do not appear to integrate HPO and public health infrastructure.
5. **Establish the role of Population Health Directors or equivalent at each HPO, who should oversee both outpatient and inpatient populations and be partially funded by public health departments**, commensurate with their role in ensuring that timely, quality-controlled data related to public health is accessible to local public health authorities. This director should be held accountable for integration of HPO information into public health information systems. While this will require an increase in funding for HPOs to support public health activities, this will reduce duplication and waste that happens in an uncoordinated set of initiatives, and savings that should be applied towards the Population Health Director position and associated resources.

Conclusion

Developing, implementing and evaluating a practical convergence plan for EHR-based data sharing networks and platforms and public health information systems requires the orchestration of expertise from several specialties. It requires public support that moves politicians to write legislation that allocates needed resources and holds recipients accountable. A careful and coordinated approach will generate consistent understanding of what is possible to be answered with EHRs and allow stakeholders to feel more confident about emerging analyses. This will also allow HPOs to focus on reducing the impact of the pandemic, rather than dealing with the multiple regulatory, logistic, and technical requests for their data. Effective data systems will enable implementation of new discoveries in communities, health systems, and policy; the implementations will directly impact the health of communities.

Unless COVID-19 data initiatives are coordinated and systems are interoperable, much effort and money will be spent into each initiative individually: these initiatives will compete with each other, will only provide partial answers, and will still not properly support public health decision making for this and the next pandemic, and for other diseases that have a large national impact. Worse, if funds are put into the development of yet another registry or COVID-19 data network, resources will be further diluted. If new infrastructure is built now to fill immediate needs for COVID-19 data, it may not create durable assets that will help us respond to the next pandemic. The public health system was caught off-guard by COVID-19, but not to be prepared for its next wave or for the next pandemic is inexcusable. Our call to action for durable assets should reduce the costs for the nation in the long-term, both in terms of direct expenditures and in personnel hours; more, it should reduce opportunity costs for organizations, where focusing on the variety of data sharing activities requested precludes these entities from implementing enhanced quality control and innovation. However, measuring impacts can be challenging. How many lives have been directly or indirectly affected is difficult to assess. If we are successful in converging valuable initiatives to the point where data from HPOs can be efficiently used in public health information systems, we should see clear differences in response to future pandemics and to endemic issues related to the exchange and analysis of health data. The return on investment (i.e., value) can be defined in terms of costs and benefits -- increased discovery of knowledge and its application to patients and health systems. In a pandemic, many stakeholders (policy-makers, health systems, public health, and, most importantly, individuals) must adapt rapidly; the cost of this adaptation in our current fragmented system has been overwhelming.

We call all stakeholders to act now to build a coordinated system of data sharing to combat COVID-19, and to prepare for the inevitable next pandemic. Successful implementation of the measures outlined in this article will enable evidence-based approaches to coordinate testing and contact tracing, predict needed resources and prepare accordingly (so “non-essential” healthcare services will not need to be shut down unnecessarily), conduct basic, preventive or therapeutic research, and provide a trusted, factual basis for answering public health questions of critical importance for this pandemic and other health conditions.

References

1. McCaw JM, Glass K, Mercer GN, McVernon J. Pandemic controllability: a concept to guide a proportionate and flexible operational response to future influenza pandemics. *J Public Health* . 2014;36: 5–12.
2. Holmgren AJ, Apathy NC, Adler-Milstein J. Barriers to Hospital Electronic Public Health Reporting and Implications for the COVID-19 Pandemic. *J Am Med Inform Assoc*. 2020. doi:10.1093/jamia/ocaa112
3. CDC. How to Report COVID-19 Laboratory Data. 1 Sep 2020 [cited 1 Sep 2020]. Available: <https://www.cdc.gov/coronavirus/2019-ncov/lab/reporting-lab-data.html>
4. Institute of Medicine (US) Roundtable on Value & Science-Driven Health Care, Olsen L, Michael McGinnis J. Organizing and Improving Data Utility. National Academies Press (US); 2010.
5. O'Reilly-Shah VN, Gentry KR, Van Cleve W, Kendale SM, Jabaley CS, Long DR. The COVID-19 Pandemic Highlights Shortcomings in US Health Care Informatics Infrastructure: A Call to Action. *Anesth Analg*. 2020. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/pmc7219836/>
6. Kannampallil TG, Foraker RE, Lai AM, Woeltje KF, Payne PRO. When past isn't a prologue: Adapting informatics practice during a pandemic. *J Am Med Inform Assoc*. 2020. Available: <https://academic.oup.com/jamia/advance-article-abstract/doi/10.1093/jamia/ocaa073/5825285>
7. Atreja A, Gordon SM, Pollock DA, Olmsted RN, Brennan PJ, Healthcare Infection Control Practices Advisory Committee. Opportunities and challenges in utilizing electronic health records for infection surveillance, prevention, and control. *Am J Infect Control*. 2008;36: S37–46.
8. Kukafka R, Ancker JS, Chan C, Chelico J, Khan S, Mortoti S, et al. Redesigning electronic health record systems to support public health. *J Biomed Inform*. 2007;40: 398–409.
9. Cifuentes M, Davis M, Fernald D, Gunn R, Dickinson P, Cohen DJ. Electronic Health Record Challenges, Workarounds, and Solutions Observed in Practices Integrating Behavioral Health and Primary Care. *The Journal of the American Board of Family Medicine*. 2015. pp. S63–S72. doi:10.3122/jabfm.2015.s1.150133
10. Committee on Review Data Systems for Monitoring HIV Care, Board on Population Health and Public Health Practice, Institute of Medicine. Monitoring HIV Care in the United States: Indicators and Data Systems. National Academies Press; 2012.
11. Brat GA, Weber GM, Gehlenborg N, Avillach P, Palmer NP, Chiovato L, et al. International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium. *NPJ Digit Med*. 2020;3: 109.
12. Haendel M, Chute C, Gersing K. The National COVID Cohort Collaborative (N3C): Rationale, Design, Infrastructure, and Deployment. *J Am Med Inform Assoc*. 2020. doi:10.1093/jamia/ocaa196
13. Vaid A, Jaladanki SK, Xu J, Teng S, Kumar A, Lee S, et al. Federated Learning of Electronic Health Records Improves Mortality Prediction in Patients Hospitalized with COVID-19. *medRxiv*. 2020. doi:10.1101/2020.08.11.20172809

14. Kruse CS, Stein A, Thomas H, Kaur H. The use of Electronic Health Records to Support Population Health: A Systematic Review of the Literature. *J Med Syst.* 2018;42: 214.
15. Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F, et al. Electronic health records to facilitate clinical research. *Clin Res Cardiol.* 2017;106: 1–9.
16. Center for Drug Evaluation, Research. COVID-19 RWD Data Elements Harmonization. 7 Jun 2020 [cited 1 Sep 2020]. Available: <https://www.fda.gov/drugs/coronavirus-covid-19-drugs/covid-19-real-world-data-rwd-data-elements-harmonization-project>
17. Califf RM, Hernandez AF, Landray M. Weighing the Benefits and Risks of Proliferating Observational Treatment Assessments: Observational Cacophony, Randomized Harmony. *JAMA.* 2020. doi:10.1001/jama.2020.13319
18. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PRO, Bernstam EV, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care.* 2013;51: S30–7.
19. Weiner MG, Embi PJ. Toward reuse of clinical data for research and quality improvement: the end of the beginning? *Annals of internal medicine.* 2009. pp. 359–360.
20. Reducing COVID-19 Deaths In Nursing Homes: Call To Action. [cited 1 Sep 2020]. Available: <https://www.healthaffairs.org/doi/10.1377/hblog20200522.474405/full/>
21. Information for Health: A Strategy for Building the National Health Information Infrastructure. In: ASPE [Internet]. 28 Nov 2016 [cited 21 Jun 2020]. Available: <https://aspe.hhs.gov/report/information-health-strategy-building-national-health-information-infrastructure>
22. Sittig DF, Singh H. COVID-19 and the Need for a National Health Information Technology Infrastructure. *JAMA.* 2020. doi:10.1001/jama.2020.7239
23. Accelerating Data Infrastructure For COVID-19 Surveillance And Management. [cited 1 Sep 2020]. Available: <https://www.healthaffairs.org/doi/10.1377/hblog20200413.644614/full/>
24. Reeves JJ, Hollandsworth HM, Torriani FJ, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. *J Am Med Inform Assoc.* 2020;27: 853–859.
25. Office for Civil Rights (OCR). HIPAA and COVID-19. 2020 [cited 19 Jun 2020]. Available: <https://www.hhs.gov/hipaa/for-professionals/special-topics/hipaa-covid19/index.html>
26. Lenert L, McSwain BY. Balancing health privacy, health information exchange, and research in the context of the COVID-19 pandemic. *J Am Med Inform Assoc.* 2020;27: 963–966.

Figure legends

Figure 1. COVID-19 data initiatives. Initiatives are categorized by data type, public/private access, individual- or aggregate-level data. Inventory resources classified as *Individual* indicate that case-level data (Protected Health Information, limited data sets or HIPAA “de-identified” data) is available to users of the resource. Resources classified as *Aggregate* indicate that summaries and averages are available to users of the resource. Bubble size = estimated size of available data. The colors indicate the types of data available to users of the resource. Only resources with a website and contact information are included (see **Table 2** for URLs of individual resources and the scale of each resource).

Figure 2. Conceptual model for an evolving digital public health ecosystem. A durable information infrastructure to overcome existing challenges requires careful coordination and leverage of existing resources. Most of the components for a future, integrated system are already in place, with the pathways moving forward requiring agreement or translation among standards, governance structures, and clear definitions of roles and responsibilities. “Chronic” public health issues refer to long-term challenges such as healthcare-acquired infections or prescription overdose. API: Application Programming Interface; CDM: Common Data Model; HIE: Health Information Exchange.

Competing Interests

SM was funded by NIH grants NCATS UL1-TR001409, NCI CA51008, NCI U24CA237719-01 and NHGRI U01 HG007437. LO-M was funded by NIH grants NIH R01GM118609, R01HG011066, NSF OIA-1937136 and Gordon and Betty Moore Foundation grant #9639. ISK is on the board of Inovalon. AB is a consultant or advisory board member for various companies. These funding agencies or commercial entities had no role in the design or contents of the manuscript.

Author Contributions

SM coordinated the compilation of COVID-19 resources. LO-M coordinated the writing of the manuscript. All authors participated in weekly discussions leading up to the description of challenges and recommendations provided in the article, wrote and edited portions of the manuscript.

Acknowledgements

The authors thank Shuo Wang, Research Associate at Georgetown University for her assistance with creating the illustrations for COVID resources and initiatives.

Data Availability Statement

The 3rd party data underlying this article and their respective data use policies are available through the resource links provided in *Table 2: Partial list of COVID-19 data initiatives*.

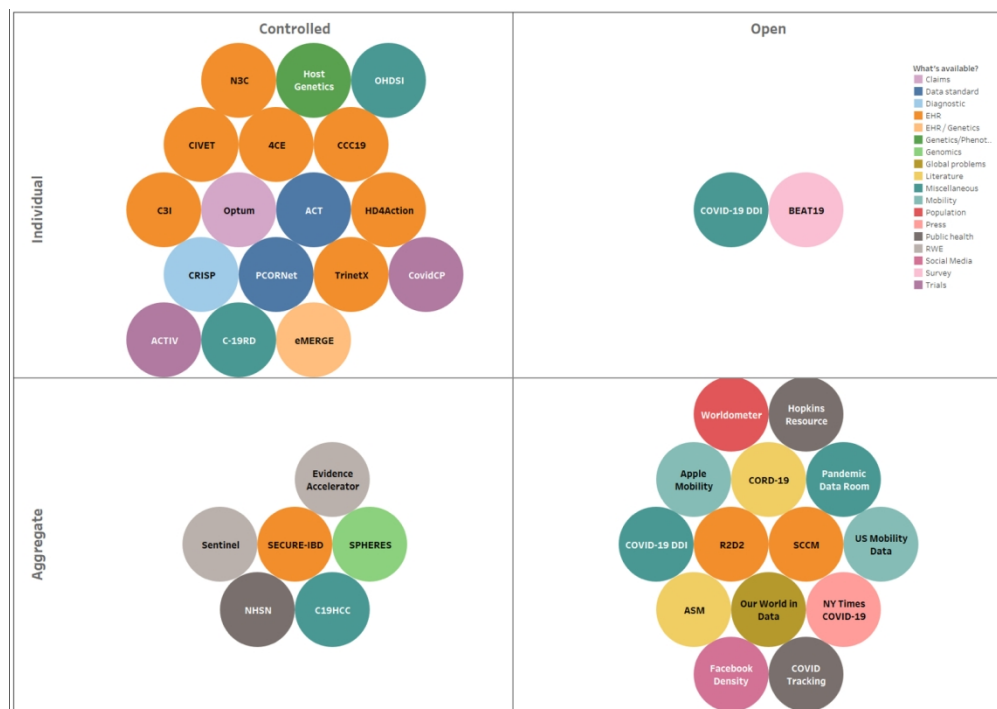


Figure 1: COVID-19 data initiatives

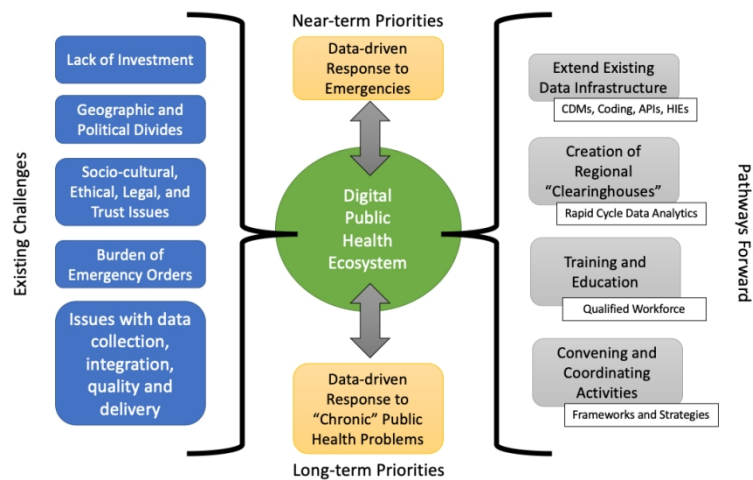


Figure 2: Conceptual model for an evolving digital public health ecosystem

338x190mm (105 x 105 DPI)

Table 1. Registries and Clinical Data Network (CDN) based on EHR derived data

Type of Data Resource	Benefits	Challenges
Disease-based (e.g., COVID-19) Registry	<ul style="list-style-type: none"> Centralized harmonization and curation of data Easier to manage Specific data items for the disease are harmonized and curated centrally Feasibility of informed consent for use of data 	<ul style="list-style-type: none"> Privacy and institutional risks associated with transferring data to a central repository Less transparency on data use Comparisons with other diseases not possible Threat of a single-point-of-failure Labor intensive if each site needs to standardize and curate the data
Clinical data network	<ul style="list-style-type: none"> Typically involves a distributed network of clinics, HPOs and/or research centers Data not restricted to patients with the disease, or to data items directly related to the disease Comparisons with “controls” and with patients with other diseases is possible No single point of failure unless there is dependency on a central hub 	<ul style="list-style-type: none"> High number of individuals and records requires additional security and privacy safeguards Detailed, curated data on the disease of interest is not always available Harmonization of complex data elements is hard to coordinate Analytics may require special methods Informed consent may not be feasible

Table 2: Partial list of COVID-19 data initiatives

Scale	Initiative	Resource name	Resource URL
Global	4CE	Consortium for Clinical Characterization of COVID-19 by EHR (4CE)	https://transmartfoundation.org/covid-19-community-project/ ; https://covidclinical.net/
	Apple Mobility	Apple Mobility Trends Reports	https://www.apple.com/covid19/mobility
	ASM	American Society of Microbiology COVID Research Registry	https://asm.org/COVID/COVID-19-Research-Registry/Home
	C-19RD	COVID-19 research database	https://covid19researchdatabase.org/
	CORD-19	COVID-19 Open Research Dataset Challenge (CORD-19)	https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge
	COVID-19 DDI	COVID-19 Data Discovery Index	www.covid19dataindex.org
	Evidence Accelerator	FDA Evidence Accelerator program	https://www.focr.org/covid19
	Facebook Density	Facebook population density	https://data.humdata.org/dataset/highresolutionpopulationdensitymaps
	Hopkins Resource Center	Johns Hopkins Coronavirus Resource Center	https://coronavirus.jhu.edu/us-map
	Host Genetics	COVID-19 Host Genetics Initiative	https://www.covid19hg.org
	NY Times COVID-19	NY Times COVID-19 Data	https://raw.githubusercontent.com/nytimes/covid-19-data
	OHDSI	OHDSI study-a-thon	https://www.ohdsi.org/covid-19-updates/
	Our World in Data	Our World in Data	https://ourworldindata.org/coronavirus-source-data
	Pandemic Data Room	Flattening the curve: COVID-19 Pandemic Data Room Visualization Challenge	https://cgdv.github.io/challenges/COVID-19/
	R2D2	Reliable Response Data Discovery for Clinical COVID-19 Consultations using Patient Observations	https://covid19questions.org
	SECURE-IBD	Surveillance, Epidemiology of Coronavirus (COVID-19) Under Research Exclusion	https://covidibd.org/
	TrinetX	TriNetX network	https://www.trinetx.com/coronavirus/
	Worldometer	Worldometer	https://www.worldometers.info/coronavirus/
National	ACT	ACT Network COVID-19: Developing COVID-19 phenotype and ontology	https://www.amia.org/sites/default/files/AMIA-COVID19-Webinar-Series-ACT-Network-CRI-2.pdf
	ACTIV	ACTIV (Accelerating COVID-19 Therapeutic Interventions and Vaccines): NIH clinical trials network for COVID vaccines testing	https://www.nih.gov/research-training/medical-research-initiatives/activ
	BEAT19	Behavior, Environment And Treatments for Covid-19 (BEAT19)	https://github.com/beat19-org/beat19-public-data
	C19HCC	COVID-19 Healthcare Coalition (C19HCC)	https://mcovid.org/
	C3I	Cancer Center Cessation Initiative (C3I) + COVID	https://sites.google.com/wisc.edu/c3icovidsmoking/grantee-resources
	CCC19	COVID Cancer Consortium (CCC19)	https://ccc19.org/
	CIVET	North American AIDS Cohort Collaboration on Research and Design (NA-ACCORD) Corona Infectious Virus Epidemiology Team (CIVET)	https://naaccord.org/covid-19
	COVID Tracking	The COVID Tracking Project	https://covidtracking.com/
	CovidCP	CovidCP clinical trials registry	-
	eMERGE	eMERGE network to support COVID research	https://emerge-network.org
	HD4Action	RWJF Health Data 4 Action COVID-19 Registry (with AcademyHealth, Health Care Cost Institute (HCCI), CareJourney, and numerous health systems)	https://www.academyhealth.org/blog/2020-04/new-initiative-aims-build-model-open-covid-19-patient-data-registry-network

	N3C	N3C (National COVID Cohort Collaborative): Building a Nationwide COVID-19 Cohort Through Informatics	covid.cd2h.org/N3C
	NHSN	CDC's National Healthcare Safety Network (NHSN) COVID-19 Module	https://www.cdc.gov/nhsn/acute-care-hospital/covid19/index.html
	Optum	Optum	https://www.optum.com/campaign/lscb/covid-19-data-dashboard.html
	PCORNet	PCORNet Mini/Thin CDM: Stand-alone, ancillary COVID-19 version of the CDM	https://pcornet.org/news/pcornet-covid-19-common-data-model-launched-enabling-rapid-capture-of-insights/
	SCCM	Society of Critical Care Medicine: Discovery VIRUS COVID-19 Registry	https://www.sccm.org/Research/Research/Discovery-Research-Network/VIRUS-COVID-19-Registry
	Sentinel	FDA Sentinel	https://www.sentinelinitiative.org/drugs/fda-sentinel-system-coronavirus-covid-19-activities
	SPHERES	CDC's SARS-CoV-2 Sequencing for Public Health Emergency Response, Epidemiology, and Surveillance	https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/spheres.html
	US Mobility Data	US Mobility data	https://docs.google.com/forms/d/e/1FAIpQLSc501xfAzEPADOWRmsdHmu-v8aN14jnKHBmEmdJJcTgRLddqw/viewform
Regional	CRISP	COVID-CRISP registry - Diagnostic tests	https://www.crisphealth.org/guidance/providers/